



Project Document Cover Sheet

Project Information			
Project Acronym	DataShare		
Project Title	DISC-UK DataShare		
Start Date	March 2007	End Date	March 2009
Lead Institution	EDINA, University of Edinburgh		
Project Director	Peter Burnhill, Mark Brown		
Project Manager & contact details	Robin Rice, EDINA and Edinburgh University Data Library Main Library Bldg., George Square, Edinburgh EH8 9LJ R.Rice@ed.ac.uk , 0131 651 1431		
Partner Institutions	Universities of Edinburgh, Oxford, Southampton and London School of Economics		
Project Web URL	http://www.disc-uk.org/datashare.html		
Programme Name (and number)	<i>JISC Repositories and Preservation Programme: Repositories Start-up and Enhancement projects strand</i>		
Programme Manager	Andrew McGregor (Amber Thomas)		

Document Name			
Document Title	<i>DISC-UK DataShare: State-of-the-Art Review</i>		
Author(s) & project role	Harry Gibbs, hgibbs@soton.ac.uk , University of Southampton, Project Officer		
Date	31 August 2007	Filename	state-of-the-art-review.pdf
URL	http://www.disc-uk.org/publications.html and http://www.disc-uk.org/deliverables.html		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1	24 September 2007	

DISC-UK DataShare: State-of-the-Art Review

by Harry Gibbs

September, 2007

Table of Contents

Introduction.....	4
Purpose of the Review	4
Scope of the Review	4
Methodology	4
Related Projects.....	5
StORe	5
RepoMMan	5
GRADE	5
CLADDIER	6
eBank UK	6
SPECTRa	6
R4L	6
IRs as Data Repositories.....	7
Researchers' Viewpoints	8
Data Produced by Social Science Researchers	9
Fig. 1: Kinds of electronic data produced by StORe respondents	9
Fig. 2: Formats of data produced by StORe respondents	10
Current Requirements for Data Deposit and Data Sharing	10
Guiding Principles for Data Sharing	11
Funders' Policies on Data Sharing	12
Table 1: Data Sharing Positions of Research Funders	13
Enforcement of Funders' Policies	15
Publishers as Enablers of Data Sharing	15
Institutional Policy	16

Current Storage/Deposit of Research Data	16
Storage of work-in-progress	17
Current Data Sharing Practices.....	17
Linking Source Data to Publication Output Repositories	18
Copyright and Intellectual Property Rights (IPR)	19
Benefits of and Barriers to Sharing.....	20
Benefits of Sharing	21
Barriers to Sharing	22
Fig. 4: Reasons that discourage data sharing amongst StORe survey respondents	23
Overcoming Barriers	23
Understanding Needs	23
Embargo Facilities	24
Knowledge Gaps	24
References	26

Introduction

DataShare is a collaborative project led by the University of Edinburgh, with the University of Oxford, the London School of Economics (LSE) and the University of Southampton. Its central aim is to develop a model for the deposit of social science datasets in institutional repositories (IRs). Lyon (2007, p.45) notes that, whilst many institutions have developed IRs over the last few years to store and disseminate their published research outputs, "...there is currently no equivalent drive to manage primary data in a co-ordinated manner."

As this review will show, although policies and practices currently operate to gather, store and preserve data, chiefly in national, subject-based data centres, much data remains unarchived and is at serious risk of being lost. DataShare believes that IRs may be developed to rescue some of this 'orphaned data' and make it available for future research, to the benefit of research communities and wider society.

Purpose of the Review

This Review has been undertaken to provide background information to inform the work of DataShare, to summarise and consolidate recent research and current policy relating to data sharing, and to identify knowledge gaps that may need to be addressed during the course of the project. It is also intended to inform the wider community, particularly librarians, of the current state-of-play.

Scope of the Review

The focus of the Review is generally restricted to the UK research environment and particular attention is paid to the social sciences. The following are addressed:

- Projects with related aims from which lessons may be drawn.
- Attitudes towards IRs as repositories for data.
- The quantity and qualities of data produced by social scientists in the UK.
- Current requirements for data deposit and data sharing, with respect to funders, publishers and institutions.
- The current storage/deposit of research data.
- Current methods of sharing data.
- The linking of source data to output publications.
- The situation regarding Intellectual Property Rights and issues arising from it.
- The benefits of and barriers to data sharing.

Methodology

This Review is largely based on reports, previous research and the experiences gained by other JISC funded projects, supplemented by responses to email enquiries, where additional information was sought. In small part, and as indicated, it draws on views expressed by social science researchers in the partner institutions. It also refers to the discussions which took place during a meeting between the author and Dr. Simon Coles in August 2007. Simon Coles is based at the School of Chemistry, University of Southampton. He has been involved in the eBank project and is currently Project

Manager of R4L. He has also gained considerable relevant experience from work on a number of EPSRC funded projects and as Manager of the National Crystallography Service.

Related Projects

This section briefly introduces a number of previously funded repository projects, focussing on the aspects of their work which bear particular relevance to DataShare¹. Several of these were part of the data cluster of the JISC Digital Repositories Programme (Jan 2005 – Nov 2007).

StORe

The central aim of *Source-to-Output Repositories* (StORe) was to develop middleware to link a source data repository, the UK Data Archive, with an output publications repository, Research Articles Online at LSE (StORe 2006). To inform this work, StORe conducted a survey of UK researchers in 2006, to investigate their opinions of source-to-output repositories and related issues. 377 questionnaires were returned from those in the disciplines of archaeology, astronomy, chemistry, biosciences, chemistry, physics and the social sciences (Pryor 2006). Of these 61 were from researchers in the social sciences. The questionnaire phase was followed by a series of individual interviews. Sixteen social sciences researchers were interviewed. StORe's deliverables include a survey report specific to the social sciences, which is of particular interest to DataShare.

RepoMMan

The *Repository Metadata and Management Project* (RepoMMan) also conducted a survey to inform its work on metadata in the University of Hull's Fedora repository. Questionnaires were distributed to researchers from a wide range of disciplines in late 2005 to gather information about research processes. The survey report (Green 2006) shows that 229 valid responses were returned, of which 59 were from the University of Hull and the remainder from elsewhere (predominantly from within the UK). The survey results concur with the findings of six preceding interviews carried out at Hull². Although the report makes no attempt to identify discipline-specific issues nor to distinguish between data and publications, it does provide some useful insights into the way in which research is carried out and stored, and into researchers' understanding of IPR.

GRADE

Geospatial Repository for Academic Deposit and Extraction (GRADE) takes a holistic approach to the technical and cultural issues around the re-use of geospatial data. Its achievements include the creation of a demonstrator repository for geospatial data³ in DSpace. It has also produced two reports that are of particular interest to

¹ A diagram showing projects affiliated with DataShare partner institutions is available at <http://www.disc-uk.org/team.html#diagram>.

² A report on the interviews was prepared in addition to the questionnaire survey report. GREEN, R. (2006) Deliverable R-D4: Report on research user requirements interview data, version 1.1. Hull: RepoMMan Project. Available from: <http://www.hull.ac.uk/esig/repomman/downloads/R-D4-rsch-int-data-11.pdf>

³ <http://gradedemo.edina.ac.uk/dspace/index.jsp>

DataShare. The first, by Robin S. Smith (2007) presents the findings of a workshop involving 13 participants from UK universities to examine various methods of sharing geospatial data. The other, by Charlotte Waelde (2007) of the School of Law, University of Edinburgh, contains complex legal discussion of IPR and digital rights under the 1996 Database Directive and draws some interesting conclusions in terms of data access (GRADE 2006).

CLADDIER

Citation, Location And Deposition in Discipline & Institutional Repositories (CLADDIER) aims to produce a demonstration system to cross-search and link publications in institutional repositories to data held in the British Atmospheric Data Centre. The demonstrator is currently available as an alpha version⁴. The work of the project includes exploration of data citation and dataset publishing (CLADDIER 2007).

eBank UK

eBank UK began in 2003. Amongst its achievements it has established an ePrints-based institutional repository in the Chemistry sub-discipline of crystallography (Duke 2007). The repository, eCrystals⁵, currently contains around 150 datasets all of which are open access. In her project evaluation report, Conole (2006) identifies further successes, including the creation of an aggregator service to search across multiple repositories and link data to output publications, and the integration of the aggregator with the RDN service, PSigate. Now in phase 3, eBank is exploring the possibility of developing an international network of crystallography repositories, termed the eCrystals Federation (Duke 2007).

SPECTRa

Submission, Preservation and Exposure of Chemistry Teaching and Research Data (SPECTRa) ran from October 2005 to March 2007 with the principal aim of developing automated tools to identify, extract and ingest data to a DSpace Institutional Repository at the University of Cambridge. Data was taken from the disciplines of synthetic organic chemistry, crystallography and computational chemistry. This successfully achieved work was informed by a survey of data handling needs in the research communities concerned. The survey included interviews and a questionnaire, and samples were taken from researchers from the project partner institutions, University of Cambridge and Imperial College London, and from the University of Southampton (SPECTRa 2006).

R4L

Repository for the Laboratory (R4L) aims to take aspects of the work of eBank a step further by setting up systems that will enable machine generated data to be semi-automatically deposited in Institutional Laboratory Repositories (Coles 2007). The scope of the project extends beyond crystallography to other areas of chemistry. Links will be made between datasets and the resultant publications and it is hoped to build a report-editing tool able to automatically produce statistics, tables and graphs to assist in the preparation of publications (R4L 2007).

⁴ <http://ignis.neodc.rl.ac.uk/claddier/search/single/>

⁵ <http://ecrystals.chem.soton.ac.uk/>

IRs as Data Repositories

The Digital Repositories Roadmap, published last year, is intended to inform JISC's future work on digital repositories and presents a vision for data access in 2010. It sees a place for IRs in filling gaps where no suitable data archive is available and states that "Institutions need to invest in research data repositories" (Heery 2006, p.16). However, Heery goes on to caution that,

No single institution is likely to have the appropriate mix of individuals to maintain and migrate for the future all the data and metadata it has produced in the previous 12 months, let alone over the institution's digital lifetime. It is therefore unlikely that departmental or institutional repositories will be the long term home of academic research data for preservation purposes (Heery 2006, p.16).

Both Lyon (2007) and RIN (2007a, p.74) found considerable unease towards IRs amongst stakeholders. RIN reports that the Research Councils tend to support subject-based data centres and many share concerns over the sustainability of IRs, fearing that

...institutional commitment might wane once the Research Assessment Exercise (RAE) is over and pump-priming funding from the Joint Information Systems Committee (JISC) comes to an end.

The Research Councils' secondary concern is that a large number of organisations all developing different systems risks resulting in inconsistency and a lack of interoperability.

In line with this, the Economic and Social Research Council (ESRC) strongly supports the use of subject data centres and continues to strengthen its relationship with the Economic and Social Data Service⁶ (ESDS). In discussions with the UK Data Archive (UKDA) Lyon (2007) found a view that little social sciences data is currently deposited in IRs because the UKDA is considered to be the 'gold star' quality service. Despite this, ESRC acknowledges that

...the growth in institutional repositories as potential recipients of data outputs, means that the landscape is more complex and further thought is required on best practice (Lyon 2007, p.18).

RIN (2007a) also found that universities themselves hold reservations about IRs in this context. In general, institutions have tended to leave data management to researchers and departments, considering it to be specialised work and correspondingly, they expect subject repositories to remain dominant, at least for the time-being. They share Heery's concern over skills, believing that more in-depth disciplinary knowledge is required in working with datasets than with publications. The fact that data management skills are called for alongside subject expertise has meant that some institutions have found it difficult to recruit suitable personnel. RIN (2007a, p.56) comments that "...career and reward structures for such data experts are not well-established".

⁶ ESDS was established in 2003, is funded jointly by ESRC and JISC, and is a collaboration between the UK Data Archive (UKDA), the Institute for Social and Economic Research (ISER), the Manchester Information and Associated Services (MIMAS) and the Cathie Marsh Centre for Census and Survey Research (CCSR).

The SPECTRa project also acknowledges the importance of subject knowledge and attributes much of its success to the discipline-specific approach, which enabled it to meet the specific needs of the research community. However, it is cautioned that their model proved highly resource intensive and, for that reason, was likely not to be widely replicated.

It is unlikely that most institutional repository managers will have either the time or the specialised expertise to engage in such narrowly-focussed work across all disciplines on a single-institution basis. Progress is thus more likely if generic solutions can be pursued and shared for implementation across multiple institutions... (Tonge and Morgan 2007, p.24).

Coles feels that an important outcome of such projects, including eBank and R4L, is that they demonstrate that 'off-the-shelf' products, such as ePrints, can be adapted to accept data, therefore eliminating the necessity for each institution to develop repositories from scratch. He envisages the eCrystals repository moving from its current position under departmental management to central Library management in the coming years. For him, the skills issue does not present a significant hurdle, in that he believes Librarians can be trained to manage data with relative ease, provided the requirements for data description are not too in-depth.

In an unforeseen move in May of this year, the Arts and Humanities Research Council (AHRC) took the decision to cease funding the Arts and Humanities Data Service (AHDS) from March 2008. The Council considers that institutions have now developed the IT skills, and have the infrastructure in place, to manage data internally.

Council believes that long term storage of digital materials and sustainability is best dealt with by an active engagement with HEIs rather than through a centralised service (AHRC 2007).

This shift of responsibility seems to conflict with the prevailing view held amongst Research Councils and would appear to be somewhat premature given current attitudes within institutions. However, it does serve to reinforce the potential role of IRs and the need for action to be taken within institutions.

The Digital Curation Centre (DCC) perceives elements of competition between subject data centres and IRs, which should be set aside to form more productive, collaborative relationships. Data centres contain essential subject and data management expertise, whilst IRs are often sited within libraries, or information services, which possess a "persistent service orientation". These respective strengths could be utilised and shared to mutual benefit (Lyon 2007, p.42).

Researchers' Viewpoints

The GRADE workshop, found that participants reacted positively to the idea of a national geospatial repository and a portal for locating geospatial datasets, but were less keen on local or departmental repositories, however no explanation for this was offered (Smith 2007).

SPECTRa found that there was generally more trust in departmental repositories than in centralised institutional services, and as a result Tonge and Morgan (2007) conclude

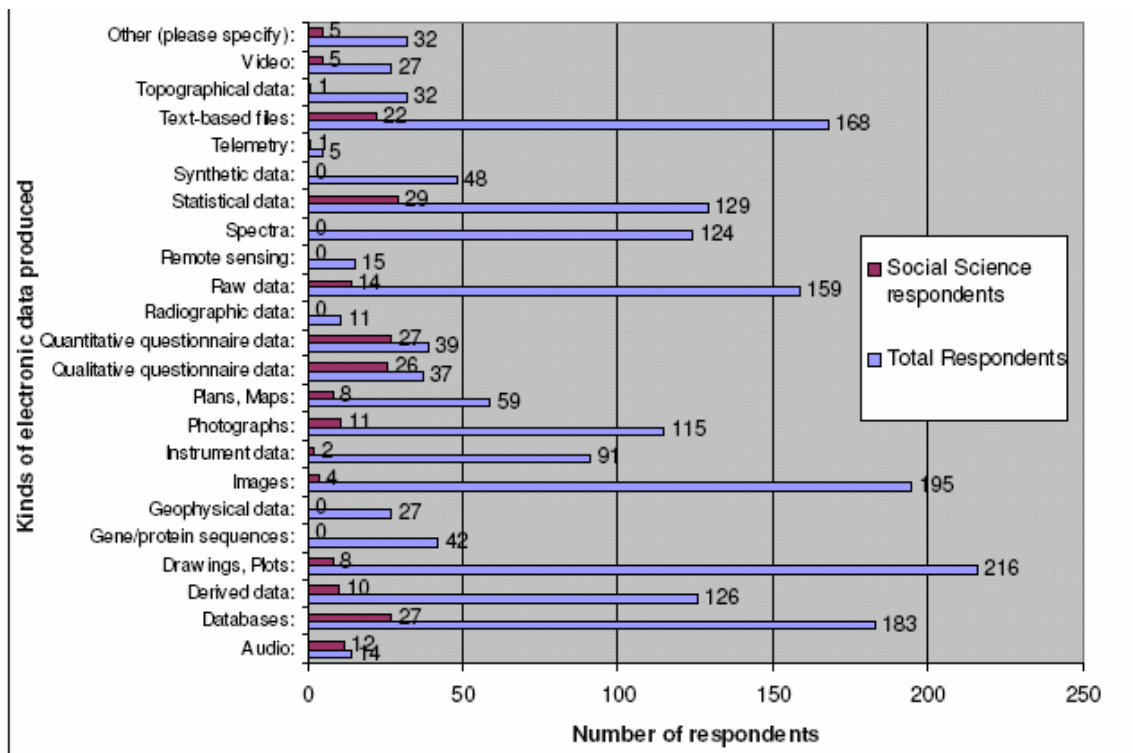
that it may be beneficial to keep data temporarily within the department during the embargo period and transfer it to the IR once it can be made more widely available.

Data Produced by Social Science Researchers

The StORe survey provides specific information about the data collected or generated by social science researchers in the UK.

Burton (2006) reports that nearly half of respondents produce quantitative data and over a third produce qualitative data, suggesting that there is a large quantity of social science data produced. Of this data, the most commonly produced types were described as statistical data, databases, quantitative questionnaire data, qualitative questionnaire data and text-based files. Other data types given included multi-media, images, maps, raw data and derived data. Figure 1 shows data types produced by social science respondents with that produced by respondents from all disciplines.

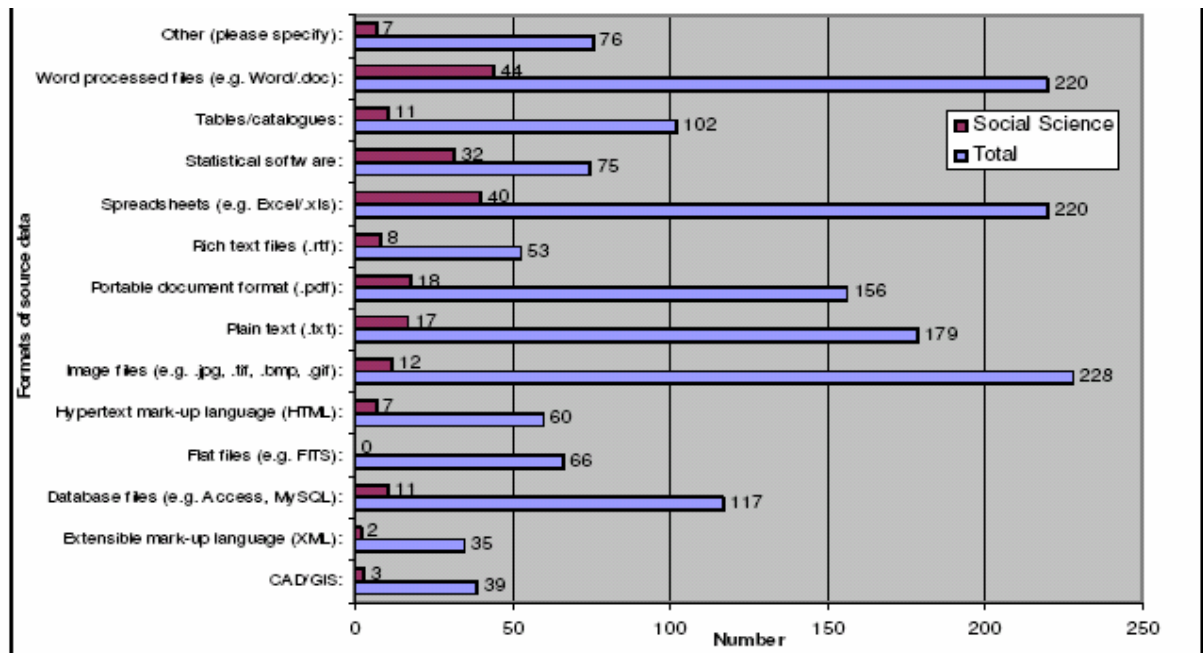
Fig. 1: Kinds of electronic data produced by StORe respondents



Source: Burton, 2006, p.19

Burton (2006) also found that the most frequently used file types are word-processed files and spreadsheet files, followed by statistical software files, although a range of other file types was given (see figure 2 for details). It is important to note that 67% of social science respondents said that the data they produce is 'often' or 'sometimes' in a combination of formats.

Fig. 2: Formats of data produced by StORe respondents



Source: Burton, 2006, p.20

Although the results of the StORe survey offer a useful indication of the quantity and type of data produced by social scientists, it does not necessarily provide the level of detail required by institutional repository managers. For instance, precise file types are not always revealed (e.g. SPSS/SAS/others in the case of statistical software), it is not known if files are compressed or uncompressed and there is no indication of the size of files, which is crucial in terms of providing adequate storage space.

It may also be necessary for an individual institution to understand in some detail what data is being produced by its own researchers, in order to accurately inform repository development. Mackenzie Smith recognised the need for work in this area at the *1st International Digital Curation Conference* in September 2005:

Discover who on campus has research data, in what forms, with what use requirements, and with what current options for archiving (Smith 2005, slide 18).

Dr. Liz Lyon reinforced this by making the following recommendation in her recent report to JISC.

JISC should develop a Data Audit Framework to enable all universities and colleges to carry out an audit of departmental data collections, awareness, policies and practice for data curation and preservation (Lyon 2007, p.45).

Current Requirements for Data Deposit and Data Sharing

The policies and practices of research funders and, more recently, publishers have significant influence over the extent to which data is shared and the methods by which sharing occurs. As such, they are addressed in this section.

Guiding Principles for Data Sharing

In 2004, in recognition of the importance of data sharing to the public good, Science and Technology Ministers requested that the Organisation for Economic Cooperation and Development (OECD) produce guidelines to facilitate access to publicly funded research data. This work was completed earlier this year with the publication of *OECD principles and guidelines for access to research data from public funding*. The document presents 13 principles which could be briefly summed up as follows. Research data should be made widely and cost-effectively available for the long-term, whilst meeting existing legal, ethical and professional requirements; emphasis should be placed on the quality and integrity of data, on ensuring that data is secured against loss or corruption, and on adhering to internationally recognised interoperability standards (OECD 2007).

The principles are attached to an OECD Recommendation which, whilst possessing 'great moral force', is not legally binding in member countries (OECD 2007). Moreover, as Rusbridge (2007) notes, the wording of the Recommendation states only that "...member states should take into consideration the Principles and guidelines...". Despite this, it may be that the impact of the document on the communities involved will help to instigate a change in culture and practice, or lead to the establishment of mandates for deposit.

In 2005, Research Councils UK (RCUK) issued a *Position statement on access to research outputs*, intended to provide a broad framework to enable more detailed guidelines to be drawn up. Although the statement focuses on publications it does recognise the value of 'underpinning data' and states "...data underpinning the published results of publicly-funded research should be made available as widely and rapidly as possible" (RCUK 2005). In its updated position statement, issued in June 2006, RCUK identified the impact of author-pays publishing and self-archiving as an area for further investigation. It is planned that a further review of the position statement will take place in 2008, when the results of this investigation are available (RCUK 2006).

In April this year, the Research Information Network⁷ (RIN) released its draft, *Stewardship of digital research data: a framework of principles and guidelines* for consultation. The paper was prepared in consultation with key stakeholders and in consideration of the OECD principles and the RCUK position statement. The five proposed principles are summarised as follows (RIN 2007b):

The roles and responsibilities of researchers, institutions and funders should be defined.

Data should be created and collected according to international standards and to a high standard of quality.

Data should be easy to retrieve and use, whilst protecting the rights of creators.

Data management should be efficient and cost effective.

Useful data should be preserved over the long-term.

⁷ RIN was established in 2005 by a UK consortium comprising the HE funding bodies, the National Libraries and the seven Research Councils. It aims to assist UK researchers by leading and co-ordinating developments in the provision of research information.

Funders' Policies on Data Sharing

Despite the guidance provided by the documents outlined above, funding organisations, such as Research Councils, Government departments and charities, do not take a uniform approach to data sharing. Although the Research Councils have all accepted the RCUK position statement, the current lack of substance might at least partially explain the significant differences between their respective policies and practices. RIN (Jan 2007a, p.55) reports that,

...the funders with the most detailed policies on data are those Research Councils which have the most developed infrastructure and resources for curating, preserving and making data accessible...

The Natural Environment Research Council (NERC) is generally considered to be one of the most advanced of the Councils, in terms of data management. NERC has had a data policy since 1996, the year in which the first of its seven data centres was established (Lyon 2007). Current data policy is contained within a detailed Handbook (NERC 2002), which is currently under review. NERC places high importance on the stewardship of data for the benefit of scientific advancement and requires that grant holders offer data to one of its data centres, where it will be maintained in the long term and made available for future research. Furthermore, NERC takes a life-cycle approach to data management in stating:

It cannot be emphasised too strongly that any proposal to undertake science which will involve the acquisition of datasets should include at the outset consideration of what is to be done with them once acquired. Valuable scientific or commercial opportunities may be lost if this fundamental principle is neglected (NERC 2002, section 5.1).

However, the current NERC policy does not advocate the concept of open data. Data is supplied at NERC's discretion and under licence terms, although data for non-commercial research incurs minimal or no charge. Commercial users, on the other hand, pay more substantial fees and Lyon (2007, p.19) notes that for the Geoscience Data Centre licensing is a "key income generator".

The Economic and Social Research Council (ESRC) also has a relatively strong record in data management, having established the UK Data Archive as far back as 1967⁸. Its current Datasets Policy, published as an annex to the *ESRC Research Funding Guide*, dates from 2000 (RIN 2007a). The policy requires that both qualitative and quantitative datasets produced in the course of ESRC funded research be offered for deposit to the ESDS/ UKDA, within three months of a grant ending. This requirement also covers data derived from ESRC funded work. Datasets should be accompanied by the necessary documentation and should be provided in a form that will enable use by third parties. Final grant payments can be withheld where these requirements are not met. Lyon (2007) notes that the ESRC's data policy is currently under review and is being widened in scope to include resources other than data under the new title 'Research Resources Policy'.

ESRC also takes a life-cycle approach to data management by encouraging grant holders to make contact with ESDS/UKDA at the earliest stage (ESRC 2007). However, much like NERC, ESDS does not make its data available on an open basis. Its remit is

⁸ It was then called the SSRC Databank; SSRC (Social Science Research Council) was the predecessor of the ESRC. The Data Archive has set up a website looking back "across the decades" at <http://www.data-archive.ac.uk/ukda40/>.

to serve the research and teaching communities, and access to data currently requires both registration and an Athens account. In addition to acquiring, disseminating and preserving data, ESDS also provides considerable support to both depositors and users of data. This includes online advice for creating and depositing data, as well as email and telephone help-lines.

Recently, both the Medical Research Council (MRC) and Biotechnology and Biomedical Sciences Research Council (BBSRC) have taken tentative steps towards open data by publishing data policies which 'expect' data to be made available 'to the scientific community with as few restrictions as possible'. The charitable Wellcome Trust has adopted a similar line for data generated under its funding.

The following table summarises the data sharing positions of the Research Councils and other funding bodies. The information is drawn from the respective policy documents and from the report by RIN (2007a).

Table 1: Data Sharing Positions of Research Funders

Funder	Data Sharing Policy	Policy Documents	Mandates/Sanctions
Natural Environment Research Council	Data must be offered to one of NERC's data centres, where it will be maintained and made available for future research.	NERC Data Policy Handbook, Version 2.2 ⁹ , 2002	Data must be offered after a 'reasonable period' of exclusive use by researcher/s.
Economic and Social Research Council	Data must be offered to ESDS/UKDA and, if accepted, will be catalogued and archived.	Datasets Policy ¹⁰ , 2000 (Annex C of <i>ESRC Research Funding Guide</i>)	Data must be offered within 3 months of a grant ending. Final grant payment can be withheld if requirements are not met.
Biotechnology and Biomedical Sciences Research Council	Data is expected to be made available to the scientific community with as few restrictions as possible in a timely and responsible manner, through existing community resources or databases where possible. Data should be available for 10 years after research ends.	BBSRC's Data Sharing Policy ¹¹ , 2007	Since 26 th April 2007 research proposals have been required to include a statement on data sharing.

⁹ <http://www.nerc.ac.uk/research/sites/data/documents/datahandbook.pdf>

¹⁰ http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Research_Funding_Guide_2005-Mar07_tcm6-9734.pdf

¹¹ <http://www.bbsrc.ac.uk/support/guidelines/datasharing/context.html>

Arts and Humanities Research Council	Following the decision to cease funding AHDS, grant holders were freed of obligations to deposit with the AHDS (except in Archaeology where a data service will be retained). Data must now be deposited with an accessible depository for at least three years after the end of a grant.	RG 28: Deposit of resources or datasets, 2007 (under AHRC Annexes to the Terms and Conditions of Research Grants in <i>Research Funding Guide 2007/08</i> ¹²⁾	Data 'must' be deposited but no time scale is prescribed.
Medical Research Council	Valuable data is expected to be made available to the scientific community with as few restrictions as possible, in a timely and responsible manner. Data must be properly curated throughout its life-cycle. MRC is participating in UK PubMed Central, which will provide facilities for data deposit.	MRC Policy on Data Sharing and Preservation ¹³ , undated	Since 1 st January 2006 funding proposals have been required to include a strategy for data preservation and sharing.
Science and Technology Facilities Council (formerly PPARC & CCLRC)	STFC has yet to develop a new data policy but PPARC's Data Curation Policy currently applies. This requires that data be made available 'to all', unless explicit reasons for not doing so are provided.	PPARC Data Curation Policy v.2, undated	Project proposals must include an agreement on data ownership, and plans for dissemination and long-term curation.
Engineering and Physical Sciences Research Council	Sees little need for data sharing within its disciplines but encourages good practice in data management.	None	None
Wellcome Trust	It is expected that researchers maximise the availability of research data with as few restrictions as possible. It is considered good practice to address data management and data sharing issues at research proposal stage.	Wellcome Trust Policy on Data Management and Sharing ¹⁴ , 2007	In some cases applicants are required to produce a data management and sharing plan and these are reviewed as an integral part of the funding decision.

¹² http://www.ahrc.ac.uk/images/research_funding_guide_2007.doc#_Toc138657015

¹³ <http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc002551.pdf>

¹⁴ http://www.wellcome.ac.uk/doc_wtx035043.html

Enforcement of Funders' Policies

In the course of this research, little evidence of the precise extent to which funders' policies on data sharing are being enforced was found. The RIN (Jan 2007a, p.56) report suggests that such information is largely unavailable: "The Research Councils that have detailed policies on data seek to secure compliance... but there is little information on compliance levels."

The impression given is that rather than concealing compliance levels, funders have simply not been in the habit of monitoring them and therefore detailed information is not available. For example, Lyon (2007, p.20) notes with regard to NERC that, "Compliance is not formally overseen and it is not clear what percentage of funded data are actually deposited in NERC data centres". Nevertheless, NERC feels that researchers tend to be keen to work with data centres and believes that the new generation of younger researchers is more aware of data issues.

In the case of ESRC, it is known that communication between the Council and ESDS/UKDA has been insufficient to ensure complete compliance in the past. This issue has been addressed, however and ESDS' 2005/06 annual report states

The ESDS support role for ESRC grant applicants and grant holders was strengthened by the development of combined record-keeping with the ESRC allowing joint monitoring of awards that result in data for deposit.
(ESDS 2006, p.6).

In response to questions posed by email, ESDS reported that this new approach has both increased the quantity of datasets offered and improved their quality due to the fact that earlier contact is being made with grant-holders¹⁵.

It is also worth noting that ESDS/UKDA is not obliged to accept data offered by researchers, although this policy may be in flux. An economist who responded to the StORe survey, for example, said that her data had been rejected on the grounds of being too narrow and time specific (Burton 2006).

Publishers as Enablers of Data Sharing

Coles believes that research publishers have as much of a role to play in facilitating sharing as funders. He noted that within his discipline, Chemistry, journal articles are the only respected form of publication and therefore journal publishers hold significant power to influence practice. Recently a requirement to provide supporting data with articles submitted to the Royal Society of Chemistry's (RSC) flagship journal, *Chemical Communications*, was introduced. Their submission guidelines state that "Experimental information must be provided to enable other researchers to reproduce accurately the work" (RSC 2007).

The eBank project team agree that publishers can play a key role but take the RSC's approach a step further in suggesting that publishers should be encouraged to access data from repositories, rather than receiving data directly (Conole 2006). This would

¹⁵ Email from Julie Missen to Harry Gibbs, 16 August 2007.

doubtless reduce the publishers' workload and may even speed the publication process to universal benefit.

It is recognised that data sharing would be greatly encouraged if data generation was recognised as a valid research output in its own right and rewarded according (Conole 2006; RIN 2007). In response to this, one of the NERC data centres and others are experimenting with the introduction of 'data journals' to enable data citation and encourage good data management through the current OJIMS project¹⁶.

Institutional Policy

Lyon (2007) identifies that institutions currently lack policy on the management, sharing and preservation of data. As such she makes the following recommendation.

Each higher education institution should implement an institutional Data Management, Preservation and Sharing Policy, which recommends data deposit in an appropriate open access data repository and/or data centre where these exist (p.47).

Current Storage/Deposit of Research Data

The inconsistencies of funders' approaches to data deposit requirements and the apparent lack of enforcement would appear to be reflected in the results of the StORe survey. Burton (2006) shows that 59% of respondents from all disciplines had never deposited data in a data repository and this figure rises to 75% in the case of social science researchers. Of the social scientists, 13% said they had deposited with the UK Data Archive (UKDA) and 16% claimed to have deposited elsewhere, including journal and project websites.

Only 4 of 61 respondents said they had deposited with UKDA more than once and the same number said they had deposited data in another repository more than once. Burton (2006) also notes that economists and sociologists are more likely to deposit with UKDA than those from other social science subjects.

Burton (2006) indicates that the majority of data is stored either on private networks or on standalone computers, whilst Green (2006), in the RepoMMan survey report, makes a number of references to the rise of USB sticks as tools for storing information.

Burton (2006) concludes that

Social science researchers give the impression of being largely self-sufficient with regard to the production of data and its storage... keeping [data] on a range of systems that are generally not open-access, from personal computers to networks based at their institutions.

¹⁶ Overlay Journal Infrastructure for Meteorological Sciences (OJIMS).
http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/repositories_sue/ojims.aspx

Storage of work-in-progress

There has been some debate within DataShare as to whether institutional repositories should store data as work-in-progress (i.e. as a number of different versions and, potentially, with multiple researchers contributing to it) or whether they should only accept final versions. Recent research provides some contribution to this discussion.

Treloar and Groenewegen (2007) report on the Australian experience of institutional repository development through three Government funded projects. Although it is still early days, they note that initial work has led to a move away from the idea of a single, all-purpose repository. Instead, they are now considering the implementation of two distinct repositories, one to enable collaboration on work-in-progress and the other to store published outputs and datasets.

The RepoMMan survey addresses the use of repositories for collaboration on work-in-progress and it is concluded that the following are necessary for a successful system: it must be possible for file access to be restricted to a closed group, a facility to prevent conflicting revisions should be available, and "it must be easy to version a file *and* to revert to an earlier version" (Green 2006, p.17).

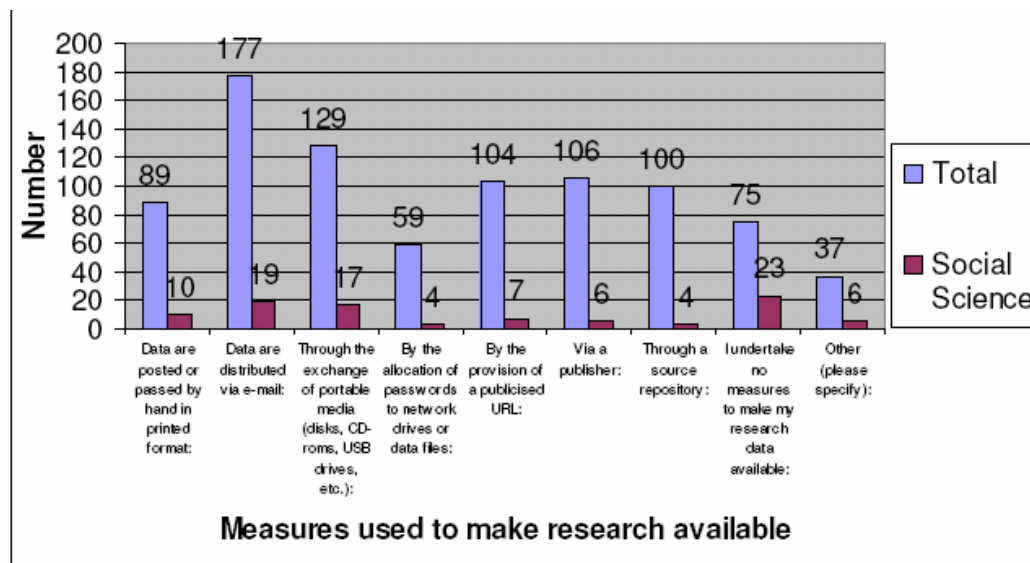
Current Data Sharing Practices

The study by RIN (2007a) found a hesitance towards data sharing amongst researchers across disciplines. In support of this, Burton (2006) reports general support for the principles of data sharing amongst social scientists; however, when asked about sharing their own data, researchers proved far more cautious. Indeed, StORe found that social scientists are significantly less likely to share their data than those from all disciplines, with 38% of respondents saying that they do not make their data available to others.

Despite this reluctance and the low level of use of source repositories, data sharing does take place, albeit in an informal manner. Findings from the GRADE study on geospatial data sharing found that sharing occurs most frequently between colleagues at the same institution and often those in the same room, although it was also found to take place across the country, worldwide and outside HE. (Smith 2007) Similarly, Burton (2006) found that sharing by social science researchers tends to result from personal requests, although the StORe survey did not specifically ask who was making these requests or how they were made (e.g. by email).

Perhaps owing to the fact that sharing tends to take place on a one-to-one basis, both studies found email and portable media, such as USB sticks, CD-ROMs and DVDs, to be commonly used. Perhaps surprisingly, 16% of StORe's social science respondents were found to share data in printed format (see figure 3), whilst Smith (2007) found that geospatial data was also often shared through private networks. Additionally, Smith's work reveals that in a small number of instances sharing is facilitated via more sophisticated technologies and Lyon (2007) notes that moves toward sharing data through social software have begun to be made in some communities, such as chemistry.

Fig. 3: StORe respondents ways of making research data available



Source: Burton, 2006, p.34

Interestingly, although only 7% of StORe's social science respondents said they make their data available through source (data) repositories, 44% had used a source repository to access others' data and this was given as the most common method of accessing data. Such high usage is reflected in the data download statistics of the ESDS annual report, e.g. over 40,000 downloads in 2005-2006 (ESDS 2006).

Linking Source Data to Publication Output Repositories

One of Burton's key conclusions is that social scientists are largely in favour of the provision of links between output publications and the underpinning source data. It may also be worth noting that postgraduate students saw significantly more advantage in this than other researchers. This is explained by the fact that established researchers generally make less use of others' research material, in terms of both publications and data. Despite this, he does not conclude that source-to-output repositories should be aimed solely at postgraduates. They may also prove useful in the process of peer-review and, in time, give rise to new ways of working, such as "analysis of research methods and publications based on the effects of this available research and data" (Burton 2006, p.55).

In their final report on the SPECTRa project, Tonge and Morgan (2007) recommend that automated mechanisms for linking between source and output repositories should be found. The related issues have, or are being, investigated by StORe, eBank, R4L and CLADDIER.

CLADDIER aims to deliver two reports which may be of use when considering the provision of links between data and publications. Respectively, these will consider recommendations for linking data to publications and users' experiences of utilising CLADDIER's demonstrator system. Unfortunately, at time of writing these had yet to be produced (CLADDIER 2007).

Copyright and Intellectual Property Rights (IPR)

The study carried out by RIN (2007a, p.14) provides a useful summary of funders' positions regarding IPR, which demonstrates inconsistencies in approach. Rights to research carried out in a university are usually held by the institution as employer, rather than the researchers, unless it is explicitly agreed otherwise. However, in practice, institutions rarely exercise these rights. Research Councils generally do not claim rights over work funded in universities, whereas Government Departments currently tend towards shared rights between the funding department and the institution. In general, RIN found that "Most funders' stance on IP issues is set firmly in the context of promoting and facilitating knowledge transfer", which is very positive in terms of repository development and the open access movement.

However, the picture regarding IPR for data would appear to be somewhat more complex. ESDS (2007) provides information and advice on copyright and IPR to potential data depositors. This includes a very useful set of guidelines for social scientists, covering copyright, database rights, confidentiality, data protection and defamation (ESDS 2000). The guide communicates some of the complexities of copyright, as regards data collection. For example, copyright in an original questionnaire is held by the creator, whilst responses to multiple-choice questions are owned by the interviewer and responses to open questions are owned by the interviewee. However, if the interviewer summarises an open response, rights to the summary are with the interviewer unless a substantial part of the respondent's words are used. Secondary use of an interviewee's response may be covered by fair dealing.

In the case of derived data, rights can be claimed by the deriver if the work can be shown to be 'original', whilst rights to the raw data remain with the original creator. Researchers may also have to consider licence terms applied to data acquired from a third party when establishing rights and ownership (ESDS 2000). JISC is currently funding a consultancy on Development IPR which is due to end in October 2007. The consultancy will provide advice on a number of areas including licensing and derived data¹⁷.

The report on IPR in geospatial databases researched for GRADE by Waelde (2007) investigates the sharing of derived data from licensed databases. She concludes that "No database copyright subsists in the structure of a geospatial database", and that "A researcher... may not be prevented from extracting a substantial part of the contents of a database for the purposes of non-commercial research... so long as the source is indicated" (p.35). She goes on to argue that researchers may legally deposit research which contains a 'substantial' part of the contents of another's 'database' in a repository, provided acknowledgement of the source is made. Likewise, substantial parts of this deposited data may then be retrieved and re-used by others for the purposes of non-commercial research.

However, Waelde's claims are disputed by the Ordnance Survey (OS). In a subsequent article in the Guardian, a spokesperson for OS commented "We haven't been able to consider [Waelde's] report in detail but there is absolutely no doubt that intellectual property rights exist in MasterMap¹⁸ - it would be ludicrous to suggest otherwise. In all

¹⁷ More information at: http://www.jisc.ac.uk/whatwedo/projects/ipr_consultancy.aspx

¹⁸ MasterMap is an online geospatial database which is provided and licensed for use by OS.

our topographic information, there is copyright as in artistic works. Therefore use of those works without licence is an infringement".

Perhaps unsurprisingly, projects which have investigated researchers' understanding of IPR issues all uncovered some element of confusion. The RepoMMan survey, for example, reports that when asked who owns the copyright to their recent research outputs, a quarter replied that they did not know. Additionally, the report suggests that the 37% who claimed to hold some degree of personal copyright were, in fact, wrong (Green 2006).

SPECTRa identifies practical issues for repositories arising from uncertainty around IPR noting,

These areas of doubt, which have important consequences for decisions about the eventual dissemination of data, need to be resolved before IPR metadata can be assigned to each data file (Tonge and Morgan 2007, p.17).

Perhaps of greater concern, GRADE found that uncertainty

...leads to a very real reluctance in making research data available for reuse. (Robertson 2006, p.2).

The Rural Economy and Land Use Programme (RELU) funded a scoping study to look at access to and management of data in the context of the programme's research activities. The study reported that the need for legal information goes beyond IPR.

...it seemed clear that respondents were in need of advice on legal issues including Freedom of Information Act and confidentiality and IP issues including data ownership" (McKay 2005, p. 19).

Both Lyon (2007) and Tonge and Morgan (2007) recommend that JISC should provide guidance for researchers on IPR issues relating to data.

Benefits of and Barriers to Sharing

Good advocacy has always been considered key to the work of DataShare and this is reinforced by the members of the eBank project team who felt that

...until researchers were aware of the possibilities and convinced of the benefits, they will be unable to take full advantage of initiatives like eBank and unwilling to change their practice (Conole 2006).

An essential role of advocates is to build knowledge of the benefits of sharing and to understand the barriers to sharing, whether these be genuine or perceived. As such, this section provides a summary of the benefits and barriers that have been identified in the course of this research and through early discussions with researchers in the partner institutions.

Benefits of Sharing

Burton (2006) reports that almost three-quarters of social science respondents to the StORe survey felt that it was 'useful or necessary' to access others' research data for their own research. This suggests that the majority of researchers in the social sciences would gain from increased data sharing and improved access to others' data.

In addition to this, the following potential benefits to individuals of depositing their data in IRs have been identified.

- IRs provide a suitable deposit environment where funders mandate that data must be made publicly available.
- Deposit in an IR provides researchers with reliable access to their own data.
- Deposit of data in an IR, in addition to publications, provides a fuller record of an individual's research.
- Metadata for discovery and harvesting increases the exposure of an individual's research within the research community.
- Where an embargo facility is available, research can be deposited and stored until the researcher is ready for the data to be shared.
- Where links are made between source data and output publications, the research process will be further eased.
- Where the institution aims to preserve access in the longer term, preservation issues become the responsibility of the institution rather than the individual.
- A respected system for time-stamping submissions (such as the 'Priority Assertion' service developed by R4L (2007)) would provide researchers with proof of the timing of their work, should this be disputed.

At institutional level, RIN (2007a) notes that there is a trend towards departments and research groups setting up websites to make their outputs available. In some cases, these may be in breach of copyright and/or may not represent the university concerned in any controlled way. The provision and use of a centralised, managed repository would overcome these issues.

On a broader level, policy makers cite cost-effectiveness as one of the key drivers for making data available (RIN 2007b; OECD 2007). Data that is shared can be fully exploited across disciplines to derive maximum value, as well as enabling funders to ensure that the same or similar research is not funded more than once. Further to this, the OECD provides a good summary of the benefits to the research community and society in general:

...improved access to, and sharing of, data: reinforces open scientific inquiry; encourages diversity of analysis and opinion; promotes new research; makes possible the testing of new or alternative hypotheses and methods of analysis; supports studies on data collection methods and measurement; facilitates the education of new researchers; enables the exploration of topics not envisioned by the initial investigators; permits the creation of new data sets when data from multiple sources are combined (OECD 2007, p.10).

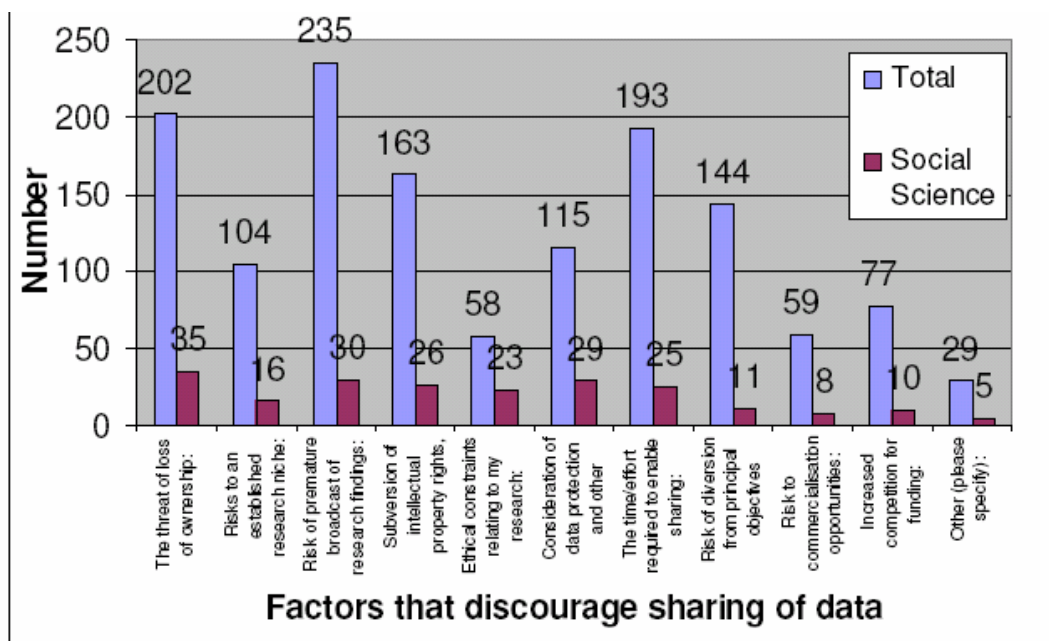
Barriers to Sharing

The following have been identified as barriers to sharing either during the course of this research or in early discussions with researchers in the project's partner institutions.

- Reluctance to forfeit valuable research time to prepare datasets for deposit, e.g. anonymisation, codebook creation
- Reluctance to forfeit valuable research time to deposit data
- Datasets perceived to be too large for deposit in IR
- Concerns over making data available to others before it has been fully exploited
- Concerns that data might be misused or misinterpreted, e.g. by non-academic users such as journalists
- Concerns over loss of ownership
- Concerns over loss of commercial or competitive advantage
- Concerns that data are not complete enough to share
- Concerns that repositories will not continue to exist over time
- Concerns that an established research niche will be compromised
- Unwillingness to change working practices
- Uncertainty about ownership of IPR
- IPR held by the funder or other third party
- Sharing prohibited by data provider as a condition of use
- The desire of individuals to hold on to their own 'intellectual property'
- Concerns over confidentiality and data protection
- Difficulties in gaining consent to release data

The results of the StORe survey provide some indication as to which issues present the greatest barriers for social scientists (see figure 4). The top two reasons for both social scientists and all scientists were found to be the threat of loss of ownership and risk of premature broadcast of research findings.

Fig. 4: Reasons that discourage data sharing amongst StORe survey respondents



Source: Burton, 2006, p.36

Overcoming Barriers

It could be argued that all that is required to facilitate data sharing is stronger mandates from funders and publishers, and more rigorous enforcement processes. However, whilst these have their part to play, experiences suggest that a more cooperative, multi-pronged approach may be more suited to the environment and, indeed, more effective. It has already been noted, for example, that data sharing could be greatly encouraged if a system of reward and recognition were introduced.

Effective advocacy has also been identified as an important factor for success and Lyon (2007) believes this needs to be accompanied by training and support, including clear guidance on IPR issues.

Understanding Needs

Coles thinks that a thorough understanding of the needs of the research community in question is central to ensuring that a repository is fit for purpose. This concurs with the findings of the SPECTRa project which placed great value on their capability to tailor the repository to fit with the workflow patterns of chemists. The project also identified a 'golden moment' which is described as

...a point at which the researcher best understands the process, possesses a comprehensive package of information to describe it, and is motivated to submit it to a data management process - which provides the optimal point for capturing data files in a repository (Tonge and Morgan 2007, p.22).

Embargo Facilities

One social scientist offered a potential solution to researchers' reluctance to share:

...entitling the producer of the research data to a 'monopoly' or 'patent' on the data for a certain period to ensure that he or she has derived maximum benefit before making it more widely available (Burton 2006, p.36).

This suggests that an embargo facility might encourage sharing, although, when asked what restrictions they would like placed on access to their data, only a small number of social scientists chose 'time related embargoes'. It is possible that researchers are distrustful of embargo facilities or the survey results may just reflect a lack of understanding of such facilities.

Interviews carried out with synthetic chemists for the SPECTRa project found strong support for an embargo facility and this, subsequently, became central to the project's data management policy. SPECTRa's self-deposit procedure requires an embargo period of 0-3 years to be set and offers the option to either 'review' or 'release' at the end of the period (Tonge and Morgan, 2007)

Knowledge Gaps

The StORe survey provides a useful indication of the quantity and qualities of social sciences data generated within institutions but it does not provide the level of detail that would be necessary if we are to ensure that 'orphaned data' is located, captured and disseminated. As has been discussed, institutions have yet to design and carry out data audits that would provide a more complete picture of the data they hold. The particulars of the number of files, file formats, file sizes, versions of files and subject areas covered have yet to be ascertained.

Similarly, previous research has suggested the need for common storage locations for data, and that much data is currently not deposited for archiving and dissemination. Still, an audit would be required to accurately discover how much data within an institution has yet to be deposited and where that data lies. It may be, however, that to undertake institutional audits is beyond the current scope of DataShare and that, as Lyon (2007) suggests, JISC should take a lead on this.

Now that we have knowledge of various funders' policies and practices, it would be useful to discover who funds social science research in the respective institutions in order to build a clearer overview of the environments in which we are working.

Whilst the experience of others has uncovered many and varied barriers to data sharing, the particular issues which concern researchers in the partner institutions remain largely unknown.

We also lack knowledge of social science researchers' preferences in terms of data sharing. Do they favour, for example, national subject-based data centres, departmental repositories or centralised institutional repositories? What are their views regarding federated search systems? Do they wish to deposit their data with their publications or would they prefer systems to link publications to sources? These

questions relate to the way in which social scientists locate and retrieve data, which should be fully understood in order to provide targeted services.

In a similar vein, it is suggested that DataShare needs to understand more about how researchers work and, perhaps, identify 'golden moments' for the social sciences, so that repository developments are fit-for-purpose and run with optimum efficiency. The data life-cycle scheme presented in the work of data archivists from the US may help the project identify the crucial points of engagement with researchers in local institutions (Green and Gutmann 2007).

References

- AHRC (2007) *AHRC news & press releases: important information for applicants to AHRC June deadline*, 14 May. Bristol: AHRC. Available from: http://www.ahrc.ac.uk/news/news_pr/2007/information_for_applicants_to_AHRC_june_deadline.asp [Accessed 26 August 2007]
- BURTON, G. (2006) *The Source to Output Repository Project: Social Science*. Available from: <http://jiscstore.jot.com/WikiHome/SocialScienceFinalReport.doc> [Accessed 10 August 2007]
- CLADDIER (2007) *Citation, Location, And Deposition in Discipline & Institutional Repositories (CLADDIER)*. [website] Chilton: BADC. Available from: <http://claddier.badc.ac.uk/trac> [Accessed 6 August 2007]
- COLES, S. (2007) *The Repository for the Laboratory (R4L) Project*. D-Lib Magazine [online], 13 (3/4). Available from: <http://www.dlib.org/dlib/march07/03inbrief.html#COLES> [Accessed 7 August 2007]
- CONOLE, G. (2006) *External evaluation of the eBank project: final report*. Bath: UKOLN. Available from: <http://www.ukoln.ac.uk/projects/ebank-uk/evaluation-report-dec-2006/evaluation-report-december-2006.pdf> [Accessed 15 August 2007]
- CROSS, M. (2007) *New study casts doubt on Ordnance Survey's copyright control* (Guardian Technology pages). The Guardian, 5 April, p.3. Available from: <http://www.guardian.co.uk/technology/2007/apr/05/freeourdata.intellectualproperty> [Accessed 6 August 2007]
- DUKE, M. (2007) *eBank UK*. Bath: UKOLN. Available from: <http://www.ukoln.ac.uk/projects/ebank-uk/> [Accessed 7 August 2007]
- ESDS (2000) *Guidelines on Copyright and Confidentiality: legal issues for social science researchers*. Colchester: ESDS. Available from: <http://www.esds.ac.uk/aandp/create/wwwcopyrightandconfidentiality.pdf> [Accessed 14 August 2007]
- ESDS (2006) *Economic and Social Data Service Annual Report 2005-2006*. Colchester: ESDS. Available from: <http://www.data-archive.ac.uk/news/publications/annrep2006.pdf> [Accessed 9 August 2007]
- ESDS (2007) *Copyright and IPR*. Colchester: ESDS. Available from: <http://www.esds.ac.uk/aandp/create/copyright.asp#a> [Accessed 14 August 2007]
- ESRC (2007) *ESRC research funding guide: post fEC*. Swindon: ESRC. Available from: http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Research_Funding_Guide_2005-Mar07_tcm6-9734.pdf [Accessed 18.06.07]
- GRADE (2006) *GRADE: Scoping a Geospatial Repository for Academic Deposit and Extraction*. Edinburgh: EDINA. Available from: <http://edina.ac.uk/projects/grade/background.html> [Accessed 24 August 2007]
- GREEN, A. and GUTMANN, M. (2007) *Building partnerships among Social Science researchers, institution-based repositories, and domain specific data archives*. Dublin, Ohio: OCLC Systems and Services: International Digital Library Perspectives. 23 (1). Preprint: <http://hdl.handle.net/2027.42/41214>

- GREEN, R. (2006) *Deliverable R-D3: Report on research user requirements on-line survey, version 1.1*. Hull: RepoMMan Project. Available from: http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research_survey_data_11.pdf [Accessed 3 August 2007]
- HEERY, R. (2006) *Digital Repositories Roadmap: looking forward*. Bath: UKOLN. Available from: <http://www.ukoln.ac.uk/repositories/publications/roadmap-200604/rep-roadmap-v15.pdf> [Accessed 15 August 2007]
- LYON, L. (2007) *Dealing with data: roles, responsibilities and relationships*, Consultancy Report. Bath: UKOLN. Available from: http://www.jisc.ac.uk/media/documents/programmes/digital_repositories/dealing_with_data_report-final.pdf [Accessed 30 July 2007]
- McKAY, H. et al (2005) *Data resources for rural sustainability research: realising their combined potential: Final report for reludata website*. York: Central Science Laboratory. Available from: <http://reludata.csl.gov.uk/final/RELU%20Final%20Report.pdf> [Accessed 8 August 2007]
- NERC (2002) *NERC data policy handbook, version 2.2*. Swindon: NERC. Available from: <http://www.nerc.ac.uk/research/sites/data/documents/datahandbook.pdf> [Accessed 10 August 2007]
- OECD (2007) *OECD Principles and guidelines for access to research data from public funding*. Paris: OECD. Available from: <http://www.oecd.org/dataoecd/9/61/38500813.pdf> [Accessed 10 August 2007]
- PRYOR, G. (2006) *Survey of researcher's use of repositories, survey report part 1: cross-disciplinary report*. Edinburgh: University of Edinburgh. Available from: <http://jiscstore.jot.com/WikiHome/SurveyPhase/SurveyFinalReport.doc> [Accessed 10 August 2007]
- R4L (2007) *R4L Project Overview*. Southampton: University of Southampton. Available from: <http://r4l.eprints.org/overview/index.htm> [Accessed 7 August 2007]
- RCUK (2005) *RCUK position statement on access to research outputs*. Swindon: RCUK. Available from: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2005statement.pdf> [Accessed 22 August 2007]
- RCUK (2006) *Research Councils UK' updated position statement on access to research outputs*. Swindon: RCUK. Available from: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf> [Accessed 22 August 2007]
- RIN (2007a) *Research funders' policies for the management of information outputs*. London: RIN. Available from: <http://www.rin.ac.uk/files/Funders'%20Policy%20&%20Practice%20%20Final%20Report.pdf> [Accessed 9 August 2007]
- RIN (2007b) *Stewardship of digital research data: a framework of principles and guidelines (Draft for consultation)*. London: RIN. Available from: <http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20-%20published%20draft%20for%20consultation.pdf> [Accessed 10 August 2007]
- ROBERTSON, A. (2006) *GRADE Progress Report, 1 June 2005 – 31 January 2006*. Edinburgh: EDINA. Available from: http://edina.ac.uk/projects/grade/public_progress_report2006-01.doc [Accessed 6 August 2007]

- RSC (2007) *Chemical Communications: Guidelines for Authors*. London: RSC. Available from: <http://www.rsc.org/Publishing/ReSource/AuthorGuidelines/JournalPolicy/CC/sect3.asp> [Accessed 22 August 2007]
- RUSBRIDGE, C. (2007) *OECD Principles and guidelines for access to research data*. Digital Curation Blog, 28 June. Available from: <http://digitalcuration.blogspot.com/2007/06/oecd-principles-and-guidelines-for.html> [Accessed 10 August 2007]
- SMITH, M. (2005) Managing MIT's Digital Research Data with Dspace. *1st International Digital Curation Conference*, Bath, 29-30 September. Edinburgh: DCC. Available from: <http://www.dcc.ac.uk/docs/dcc-2005/m-smith-dcc-2005.ppt> [Accessed 20 August 2007]
- SMITH, R. S. (2007) *Geospatial data-sharing in UK Higher Education: informal repositories and users' perspectives*. Sheffield: University of Sheffield. Available from: http://edina.ac.uk/projects/grade/Grade_reportRSSv2.pdf [Accessed 15 August 2007]
- SPECTRa (2007) *SPECTRa*. [website] Cambridge: Cambridge University Library. Available from: <http://www.lib.cam.ac.uk/spectra/index.html> [Accessed 25 August 2007]
- StORe (2006) *Source-to-Output Repositories*. [website] Available from: <http://jiscstore.jot.com/WikiHome> [Accessed 26 August 2007]
- TONGE, A. and MORGAN, P. (2007) *Project SPECTRa: JISC Final Report*. Cambridge: University of Cambridge Library. Available from: http://www.lib.cam.ac.uk/spectra/documents/SPECTRa_Final_Report_v10.doc [Accessed 7 August 2007]
- TRELOAR, A. and GROENEWEGEN, D. (2007) *ARROW, DART and ARCHER: a quiver full of research repository and related projects*. Ariadne [online] 51. Available from: <http://www.ariadne.ac.uk/issue51/treloar-groenewegen/> [Accessed 6 August 2007]
- WAEELDE, C. (2007) Geospatial databases and the research and education sector in the UK: designing a licensing strategy for sharing and re-use of data. Edinburgh: GRADE Project. Available from: <http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf> [Accessed 6 August 2007]